DOCUMENT RESUME

ED 214 379 FL 012 828

AUTHOR Wesdorp, H.
TITLE Backwash Effects of Language-Testing in Primary and Secondary Education.
PUB DATE Aug 81
NOTE 25p.; Paper presented at the World Congress of the International Association of Applied Linguistics (6th, Lund, Sweden, August 9-15, 1981).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Achievement Tests; Dutch; Elementary Secondary Education; English (Second Language); French; German; Language Tests; *Multiple Choice Tests; *Testing; *Test Validity
IDENTIFIERS Netherlands

ABSTRACT

A debate has been carried on in Dutch educational circles about the widespread use of multiple-choice tests, and a number of objections have been raised against the use of such tests. This paper reports on research into the validity of the objections, in particular with respect to the possible effect of multiple-choice tests on the teaching of Dutch as a native language, French, German, and English. The objections dealt with are the following: (1) the possible discriminatory effect of multiple-choice tests; (2) the supposed impoverishment of education through the use of such tests; (3) the risk of impoverished teaching methodology; (4) the possible loss of certain complex skills such as reading; (5) the prominent place of multiple-choice tests at the end of secondary education with the consequent possible overuse of the tests as practice material; (6) the possible influence of multiple-choice tests on the study habits of students; and (7) emotional aspects of the tests, particularly an increased fear of test taking. The overall conclusion seems to be that in the Netherlands many initial fears about multiple-choice testing were not based on facts. Furthermore, the Dutch system has continued the use of other testing methods along with the multiple-choice variety. (AMH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED214379

Stichting
Centrum voor
Onderwijsonderzoek
van de
Universiteit van Amsterdam

BACKWASH EFFECTS OF LANGUAGE-TESTING IN PRIMARY AND SECONDARY EDUCATION

Paper presented at the 6th AILA World Congress
Lund, Sweden, August 9-15, 1981

H. Wesdorp

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

H. Wesdorp

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

FL012828

Stichting Centrum voor Onderwijsonderzoek / Centre for Educational Research
Singel 138 - Amsterdam

BACKWASH EFFECTS OF LANGUAGE-TESTING IN PRIMARY AND SECONDARY EDUCATION

0. INTRODUCTION

Since the mid-nineteen-sixties, the use of objective testing in education, in particular of multiple-choice tests, has increased dramatically in the Netherlands. In 1968, the National Institute for Educational Measurement (CITO) was founded and since then the construction of many achievement tests has been concentrated there. This general changeover from traditional testing to objective testing also applies to language teaching. It has resulted in the construction of a fair number of multiple-choice tests, not only for Dutch as a mother-tongue, but also for French, German, and English, the three most important foreign languages taught in Dutch schools. Multiple-choice tests for Dutch as a mother-tongue are now part of the final testing program at the end of primary education which plays an important part in the selection for secondary education. Some types of secondary education also incorporate mulptiple-choice tests for Dutch in the final examinations which determine a student's admission to higher education. The same holds for foreign language tests: they form an important part of the final examinations in various types of secondary education.

From an educational point of view, the introduction of a new type of test for an important selection procedure is certainly not a trivial matter: the new tests are frequently feared for their possible influence on education in general. Most of the time they provoke heated discussions in journals of education, but the debate generally subsides after a while, when teachers have become used to the new tests. Eventually, the educational system accepts the new situation and the discussion stops altogether. At first glance, such a development seems perfectly acceptable, but this course of events also has its problematic sides. The developers of the new language tests are primarily interested in perfecting the technical quality of their measuring instrument and do not worry so much about its effects on the educational system. Nevertheless, these effects determine in part the validity of the newly introduced instrument. After all, tests are made to accomodate educational goals and their content-validity is defined as the relationship between the content of the test and of the educational program. If a test with only moderate content-validity is introduced in a high-status testing situation, for example in an important selection procedure, chances are that the educational system will adapt to the questions being asked in the test. As time

goes on, this process will increase the validity of the test-content and, after a while, the fit between the educational system and the test will be perfect, giving the test perfect content-validity, while the original discussion about the discrepancy between test-content and educational goals recedes into the past.

Those who construct tests should be more aware of such developments. Such an awareness should, in the first place, lead to the development of tests with a perfect fit to the existing educational goals, so that backwash effects will be out of the question. In the second place, this awareness should be an incentive for research into the possible effects of introducing the measuring-instrument in question, in case of doubt about the desired fit. In this way, test-constructors would achieve a more balanced view on the various aspects of content-validity of their testing instruments.

1. SUMMARY

The introduction of multiple-choice tests in the Netherlands in the early nineteen-sixties has been a cause for heated debate in Dutch educational circles and a number of objections have been raised against the use of such tests. The present paper reports on research into the validity of these objections, in particular with respect to the possible effect of multiple-choice tests on the teaching of Dutch as a mother-tongue, French, German, and English (Wesdorp, 1979).

The first objection concerned the possible discriminatory effect of multiple-choice tests (par. 2). Multiple-choice tests were thought to put certain social groups at a disadvantage and, consequently, an attempt was made to assess their discriminatory effect in two areas: reading comprehension in the mother-tongue and English. No discriminatory effects were demonstrated.

The second objection concerned the supposed impoverishment of education through the use of multiple-choice tests (par. 3). It was thought that skills which could not be measured through multiple-choice tests would not be practised in the schools and would, therefore, gradually be lost. An investigation of the writing skills at the end of primary school in schools

where a multiple-choice writing test had been the final test during the past twelve years showed no differences in quality between essays written in 1966 and in 1978.

The third drawback of multiple-choice tests was thought to be the risk of impoverished teaching methodology (par. 4). Supposedly, a final language examination with multiple-choice tests would preclude the use of certain teaching methods, causing a gradual decline of their use. A comparison between teacher-activities in schools with and without a multiple-choice final test did not point up any clear differences.

The fourth objection was raised against the possible loss of certain complex skills (par. 5), such as reading skills. Multiple-choice tests were said to be less capable of measuring complex skills than conventional tests, and this could have an effect on education. An analysis of reading comprehension skills in both types of mother-tongue tests showed, at most, a difference in favor of multiple-choice tests.

The fifth kind of criticism was leveled against the prominent place of multiple-choice tests in the school setting (par. 6). It was feared that multiple-choice tests would become important practice-material, as a result of the multiple-choice exam at the end of secondary education. A survey of language teachers revealed no such prominent role, although multiple-choice tests may, for some subjects, be used too frequently in class.

The sixth drawback was seen to be the possible influence of multiple-choice tests on the study-habits of students (par. 7). Students who are used to studying for a multiple-choice test were expected to prepare themselves in a different way than students used to traditional test forms. It turned out, though, that students who had been confronted relatively frequently with multiple-choice reading comprehension tests for English showed similar study-habits as students frequently tested with other test forms. Neither group did much preparation for such tests, anyway. In other subjects there did turn out to be a relationship between study-habits and test format.

The seventh type of objection concerned certain emotional aspects of multiple-choice tests (par. 8). Supposedly, multiple-choice tests would not

contribute to a more realistic self-image on the part of the student. They were also said to increase fear of test-taking and to be less 'acceptable' to students than traditional testing instruments. A survey of students revealed that, in general, they do not have a negative attitude towards multiple-choice language tests, although they clearly do object to some aspects of multiple-choice testing.

2. THE DISCRIMINATORY EFFECT OF MULTIPLE-CHOICE TESTS (READING COMPREHENSION)

From earlier surveys and from the Dutch literature on the subject, it became clear that to many people the introduction of multiple-choice tests in examinations implies an increase in selection on the basis of socio-economic status. Multiple choice tests are looked upon as undemocratic and are thought to place certain social groups or regions at a disadvantage when such tests are part of an examination.

These opinions often appear to be based on the idea that multiple choice tests appeal more to certain verbal skills than traditional instruments. This is seen to be a disadvantage for certain social groups, particularly when they are also confronted with a middle-class language, of which they do not have a good command.

In this investigation the supposedly discriminatory effect of multiple-choice tests has been assessed in two different ways. One part of the investigation was carried out in the final year of HAVO-schools, for English, while the other part was directed at the LBO-examinations for Dutch.[1] Object of the investigation was to find out whether multiple-choice tests discriminate against certain social groups, as compared to traditional tests.

For the first investigation, parallel instruments were constructed which were as much equivalent as possible. One consisted of multiple-choice questions and one of essay-questions. Two versions were made of this instrument, so that half of the students was examined first on the multiple-choice questions and then on the essay-questions, while the other half was tested in the other order. As far as content was concerned, all students answered the same sixteen questions. The tests, together with questions concerning the occupation of the student's father, were administered to 412 final year students from twelve HAVO-schools.

The average degree of difficulty (p-value) of the multiple-choice questions was .57 and of the open questions .34. The test which first presented the multiple-choice questions, followed by the essay-questions, showed slightly better results than the test presenting the questions in the opposite order. The reliability (Cronbach's α) of both test versions was .63. The average scores for students of different socio-economic status is shown in Table 1. Considering the low reliability of the tests, the differences in the scores of Table 1 should be interpreted with caution.

Table 1.

The test results have been analyzed in a number of ways. Separate analyses of variance for both test halves did not show any interaction effect between socio-economic status and question format. An analysis of score differences did not indicate a discriminatory effect for multiple-choice questions, while the relative results for boys and girls on essay and multiple-choice questions yielded no difference either. Significant differences were found between schools, but there is no question of an interaction effect between school and socio-economic status; in other words, none of the schools was found to have a socially discriminating influence.

The results of this first investigation do not in any way indicate an 'anti-democratic effect' for multiple-choice questions. It is, of course, possible that the attempt to assess the effect of the differences in question format as accurately as possible, by means of keeping the question-content constant, has removed the essential differences between the two types of questions from the experiments. The presumed differences between multiple-choice and open questions do not only consist of presence or absence of a list of alternative answers under the question. One criticism often leveled against objective tests is their poorer content as compared to traditional tests.

A comparison of traditional and multiple-choice tests --both representative for their category-- was the purpose of the second investigation. The current situation for LBO final exams offered a good opportunity for such a comparison. One track of LBO has a two-part final examination for a number of subjects, including Dutch. The students take one exam, consisting of

open questions and one consisting of multiple-choice questions. By collecting both types of results for each candidate, as well as data on socio-economic status, a comparison could be made of the difference in achievement on either type of test for students from various social groups. By using data from actual exams, maximum student motivation and achievement is guaranteed, while artifacts of an experimental situation can also be avoided.

For 258 final-year students from seven LBO-schools the final exam results for Dutch were collected in 1977 and a classification was made of the father's occupations. In Table 2 the exam results are given as follows:

- Average scores and standard deviations for the essay-questions on the examinations: maximum score 57.
- Average scores and standard deviations for the multiple-choice part of the examination: maximum score 50.

Table 2.

The table shows that the students scored higher on the essay-questions than on the multiple-choice part of the test, while the standard deviations on the essay questions are also higher. The correlation of .36 between the two formats is significant, but rather low, presumably because both test forms measure a number of skills simultaneously. Neither the open questions nor the multiple-choice questions showed significant differences, according to socio-economic status, although the averages for multiple-choice questions for higher status students were somewhat higher. The average differences of the five status groups seem to show the predicted tendency, but there was no question of any significant difference. Since the reliability of the scored differences is probably also quite low in this case, given the low reliability of the scores for both question formats, these results should be interpreted with caution. By means of an analysis of variance, a significant difference between the seven schools on both types of questions could be found, but there was no indication of an interaction effect between

school and socio-economic status. In other words, none of the schools seemed to show an influence of socio-economic status on the differences in results for the two test formats.

From this investigation it may be concluded that *a socially discriminating effect for multiple-choice tests in the 1977 final examination for Dutch in LBO-schools and in the constructed HAVO-tests for English could not be demonstrated. Achievements on multiple-choice and open-question tests for Dutch and English were not found to be different for the various participating social groups.*

3. INFLUENCE ON THE CURRICULUM: THE INFLUENCE OF MULTIPLE-CHOICE TESTS ON THE WRITING SKILLS OF THE STUDENT

In 1966, the first achievement test was administered in Amsterdam in the final year of primary education. This achievement test has slowly acquired a certain degree of status, because its results determine in part to which type of secondary education the student will be referred. The instrument consists entirely of multiple-choice tests on all the important subjects taught in primary school. For Dutch as a mother-tongue the test contains subtests on reading comprehension, spelling, grammar, and an attempt to measure writing skills in an indirect way. An essay containing various mistakes is used to assess a number of writing skills. The student is requested to correct the mistakes, choosing one from a few alternatives. The mistakes concern skills in a variety of areas. These areas are: writing conventions, i.e. dealing with problems of spelling and interpunction, syntax, appropriate and effective language-use, i.e. stylistic nuances, organization of the essay, i.e. dealing with structural mistakes, such as lack of necessary information, poor use of paragraphs, and content, i.e. mistakes in the content of the essay. The first year in which the exam was given (1966), all participants also wrote an essay, which was not evaluated in any way.

It was thought that the limited scope of the test-content (only multiple-choice items, no essay-questions, no testing of productive language skills) would lead to less attention in the schools for independent essay-writing, eventually leading to a decrease in writing skills on the part of the stu-

dents. The 1966 requirement to write an additional essay made it possible to compare writing skills in 1978 and 1966. A comparable sample of students was required to write an essay, similar to the one in 1966. The two sets of essays were then compared for quality, presupposing that the effect of twelve years of multiple-choice testing would have resulted in poorer quality of the 1978 essays.

Two comparisons were made in this investigation. First, 64 'old' and 64 'new' essays were compared for overall quality by 15 judges, giving a global judgement. Obviously, there were no differences in format between the old and the new essays and they were mixed randomly. Secondly, 32 old and 32 new essays were judged analytically: i.e. judgements on 'content', 'style', and 'writing conventions' were elicited from 16 (different) judges. Differences were expected, both in the comparison of the global judgements and of the analytic judgements, which could presumably be explained by the 'impoverishment-effect' of the achievement test. The reliability of the jury-judgements (the degree of consensus with other jury's of equal size and competence) varied from .78 to .95 for the various criteria. Table 3 shows some results of the investigation. The global and analytic total judgements were given on a ten-point scale, the type of scale generally used in schools in the Netherlands, while the analytic categories were judged on a four-point scale.

Table 3

None of the expected differences was actually found in the investigation, leading to the conclusion that *no impoverishment as a result of the use of achievement tests is to be found in the teaching of writing skills.* Enrichment in the teaching of more technical aspects of writing skills, such as 'writing conventions' could not be found either, even though the test does especially stress such points every year.

4. INFLUENCE ON THE CURRICULUM: THE INFLUENCE OF MULTIPLE-CHOICE TESTS ON TYPES OF TEACHING AND ON STRESSING CERTAIN PARTS OF THE CURRICULUM

A third investigation was designed to see whether the achievement test, with its limited coverage of the goals of mother-tongue education, had any effect on teaching practices. To this end, a comparison was made of the time spent on various parts of the curriculum and of types of teaching in 102 schools. 51 schools had been using the test for quite some time and 51 had not used the test at all. Obviously, the expected differences were that the 'test-schools' would pay more attention to components of the curriculum that are part of the test and less to components that will not be tested, in comparison with the 'non-test-schools'. Two types of information were elicited by means of a questionnaire: an assessment of the time spent on various types of teaching and of the teacher's attention for different parts of the language curriculum.

Table 4.

Table 4 shows that few differences could be found. The differences were subjected to the Mann-Whitney U-test, which showed that both oral and written quizzing is done more frequently in the 'test-schools'. Table 5 shows the way in which the teacher's attention was divided among the various parts of the language curriculum.

Table 5.

No differences could be traced in the time spent by teachers on various parts of the curriculum, even though some differences had been expected. Consequently, the investigation led to the conclusion *that the use of achievement tests has no clear influence on teaching practices.*

5. INFLUENCE ON THE CURRICULUM: A COMPARISON OF OLD AND NEW READING COMPREHENSION EXAMS FOR THE MOTHER-TONGUE AT THE END OF SECONDARY EDUCATION

For one type of secondary education, Lower Technical Schools (LTO), it was possible to compare traditional reading comprehension exams using a test with open questions, with the multiple-choice reading comprehension tests using a text followed by a number of questions. During a number of years, these two examination-formats existed side by side, which made it possible to compare the skills that were being measured by the two types of tests between 1971 and 1977. This was done by using the criteria for measuring reading comprehension skills as they have been formulated by Davis (1971):

Nine operational skills of comprehension in reading among mature readers

1. Word knowledge, as measured by recognition vocabulary items
2. Ability to select the appropriate meaning for a word or phrase in the light of its particular contextual setting
3. Ability to follow the organization of a passage and to identify antecedents and find references in the passage
4. Ability to select the main thought of a passage
5. Ability to answer questions which are answered directly in the passage
6. Ability to answer questions which are answered in the passage but not in the words in which the question is asked
7. Ability to draw inferences from the passage about the content of the passage
8. Ability to recognize the literary devices used in a passage and to apprehend its tone and mood
9. Ability to determine the writer's purpose, intent, and point of view; i.e., to draw inferences about the author

Table 6 gives the percentage for each category of the Davis system which is required for answering the two types of exam-question, according to the judgements of a three-member jury. The open-question exams contained a total of 276 questions, the multiple-choice exams contained 447 questions. The average inter-judge reliability was 57%, which is why the results for the three judges have not been collapsed into one figure. According to all three judges, there are differences in the way the skills for the two types of examinations are divided among the Davis categories.

Table 6.

The three judges found that on open-question exams, questions about word knowledge (Davis skill 1), and grammatical structure and function (skill 3) are asked more frequently. Both multiple-choice and open-question exams frequently demand answers that are found directly in the passage (skill 5). Multiple-choice questions often require the ability to select appropriate meaning from context, recognize paraphrases, or draw inferences (skills 2, 4, 6, 7 and 8). In order to see to what extent these differences are due to a general change in exams over the period from 1971 to 1977, it was decided to compare the Davis classification for the last three exams separately (two exams from 1976 and one from 1977). Table 7 shows the required skills in percentages, according to the judgements from three judges.

Table 7.

Testing the frequency-distribution of the various skills for the two exam-types yielded significant values for all three judges. According to all three judges, skills 6, 7, 8 and 9 were required more in multiple-choice tests. This means that, according to the judges, the exams in the last two years required more ability to paraphrase, draw inferences, recognize literary devices, and determine the writer's intent and point of view. In 1976 and 1977, both multiple-choice and open-question exams required less word knowledge than did earlier exams.

The conclusion that may be drawn from this investigation is that the analyzed examinations LTO-reading comprehension have not become poorer in content as a result of the introduction of multiple-choice tests. On the contrary: a certain degree of enrichment can be found in comparison with the open-question examinations. Inasmuch as the differences between the two types of exams have become smaller in recent years, or have disappeared completely, this seems to have resulted mainly from an improvement in the essay-questions after the introduction of multiple-choice questions in final exams.

6. INFLUENCE ON THE CURRICULUM: THE USE OF MULTIPLE-CHOICE TESTS IN THE CLASSROOM

The degree of test-use in the classroom could be considered an indicator for possible influence of multiple-choice final examinations on the curriculum. The more frequently multiple-choice tests are used in class, the more likely it is that teaching-content is reduced to teaching only what will be tested on the final examination. In this section a summary is given of a survey of the frequency with which multiple-choice tests are used for teaching in the final years of secondary education, as well as the other uses to which such tests are put.

A questionnaire was sent to sample schools for secondary education, in particular to the school-sections, whose subjects are partially or completely tested by means of a multiple choice test in the written final examination. 965 schools were approached, including 4471 sections. The percentage of responses was 47% (2102 questionnaires). The survey yielded a large amount of interesting data, which cannot all be summarized here, but the most striking results for language-teaching will be given below.

The first part of the survey was concerned with the format of test questions used in the classroom. In the final year, the percentage of multiple-choice questions versus essay and short-answer questions was 21 versus 15 for French, 23 versus 11 for German, and 27 versus 14 for English. For Dutch the percentages were higher: 34 versus 30. These figures imply that multiple-choice tests only have limited influence on classroom test format, while the influence is greatest for Dutch.

The second part of the survey was designed to find out whether multiple-choice tests are used as teaching aids in the classroom and as homework-assignments. The results for language teaching have been summarized in Table 8 and show that multiple-choice tests are not being used on a large scale. Special exercises to train students for multiple-choice questions are given frequently, though.

Table 8.

The third purpose of the questionnaire was to find out whether teachers recommend a special study-method for students who will be taking multiple-choice tests on a particular subject. A sizeable group, though still a minority of the teachers being questioned, does give special study recommendations (for Dutch 45%, French 38%, German 37%, and English 38%). From additional remarks made by the respondents, it is clear that these recommendations mostly concern techniques for tackling multiple-choice questions. For dealing with foreign language tests, teachers generally advise their students to read the text a number of times, look at the multiple-choice questions, find the relevant passages, review the text, and then proceed to answer the question. For tackling the questions themselves, the most frequent advice is to use the method of elimination, or to answer the question first as if it were an essay-question and then look at the alternatives. Foreign language teachers in all school-types consider increasing the student's knowledge of idioms and reading newspapers and books in the foreign language (without a dictionary) the best test-preparation.

The influence of multiple-choice tests on foreign language teaching in the Netherlands is not very substantial --according to the reported use of such tests in the classroom--, but on the other hand, it is not negligible. In any event, "a total loss of flexibility as a result of the hegemony of multiple-choice tests" is out of the question. Fortunately, a variety of tests is being used in the final years of secondary education.

7. INFLUENCE ON THE STUDENT: THE INFLUENCE OF MULTIPLE-CHOICE TESTS ON THE STUDY-HABITS OF STUDENTS

Many kinds of influence on education from tests and examinations have been studied, but the influence of tests on student's study-methods has been studied more than most. In a large body of investigations students have answered questionnaires, stating that they study differently for multiple-choice tests than they do for traditional --open-question-- exams. Most of these investigations have been directed at factual school-subjects. An earlier Dutch survey among foreign language teachers showed that 43% of the teachers perceive students' study-habits for foreign languages as being influenced unfavorably by the use of multiple-choice reading comprehension tests, both in exams and in the classroom.

In order to pay special attention to this problem, an inventory was made in 10 HAVO final-year classes of the study-methods used by students who are either frequently confronted with multiple-choice tests, or with traditional testing instruments. The teachers in the participating schools had been asked earlier, how many times per month and per year they gave their students multiple-choice tests and how often open-question tests. The frequencies reported by the teachers were used to form two groups of students for each subject, one with relatively little and one with relatively substantial experience with multiple-choice tests. After taking the tests, students were given a questionnaire about their preparation methods for open-question or multiple-choice tests for English, depending on their reported experience with multiple-choice tests. Students who did not do any preparation at all did not fill out the questionnaire. 99 students from 5 schools were given a questionnaire about multiple-choice tests, while 85 students from another 5 schools filled out a questionnaire about preparing for open-question tests. The questionnaires contained 11 questions about preparation methods for tests, one question about student preference, and one question about the type of test-questions students found most difficult.

The experiences of students with open-question tests did not always coincide with the frequency of those tests reported by the teachers. To a lesser extent, this was also the case with multiple-choice tests. Most students report little experience with open-question tests: only 17% reports more than one such test per year. Students turn out to do little or no preparatory work for English tests. For a multiple-choice test, 2% of the students studies always or most of the time, 20% sometimes, and 78% (almost) never. For an open-question test, these percentages are 32%, 20%, and 48%, respectively. 84% of the students spends an average of less than half an hour preparing for a multiple-choice test, while for an open-question test 46% prepares less than half an hour, 29% between half an hour and an hour, and 24% more than an hour. The preparation-time for open-question tests is spent mostly on learning words and expressions and looking through lists of idioms. Old reading comprehension exercises and tests are less popular as study material. The answers showed a strong student preference for multiple-choice tests (72%).

Although students still study a little longer for open-question tests, they generally do not prepare much at all for English reading comprehension tests.

Within the group with more multiple-choice experience, a strong correlation was found between the frequency of being tested on multiple-choice tests and the degree of preparation.[2]

8. INFLUENCE ON THE STUDENT: OPINIONS OF STUDENTS ABOUT MULTIPLE-CHOICE TESTS

This part of the investigation was an attempt to determine whether students' objections and complaints mentioned in the literature regarding multiple-choice tests were actually found among students in Dutch schools.

A questionnaire with 26 questions was presented to samples of approximately 200 students for the subjects Dutch, English, and German, in LTO-(lower stream of secondary education), MAVO-(middle stream) and VWO (higher stream)-schools. The questionnaire contained questions about seven types of statements that frequently come up in the literature. The results for each category of statements are given below.

Preference or dislike of multiple-choice tests. The majority of the students prefers multiple-choice to open-question tests, but this positive attitude is weakest for German. Most students (70%) would like more practice with multiple-choice tests.

Multiple-choice tests and the student's self-image. Is taking multiple-choice test the most eminent way for students to gain insight into their own capabilities? Students do not express an opinion about this aspect. About half of the students recognize the positive-feedback qualities of multiple-choice tests (for Dutch and English). For German they are more negative on this score: only a minority believes that taking multiple-choice tests tells them anything about their own level of knowledge.

Multiple-choice tests and fear of test-taking. A certain degree of test-taking anxiety is certainly produced by multiple-choice tests: 70% of the students disagrees with the statement that multiple-choice tests 'let you know where you stand'. In general, they find multiple-choice tests full of treacherous pitfalls, but only a minority finds multiple-choice questions so confusing as to lower the student's performance.

Statements about measuring-problems in multiple-choice tests. Students do not think much of the validity of reading comprehension tests: they express absolutely no faith in the test's capability to measure what they really know. About two thirds of the students objects to the phrasing of the alternatives in the tests. The performances for the other statements are in contradiction with some frequently held opinions: in general, students believe that they need to thoroughly understand the material in order to take the test and that it is not in the first place a matter of applying smart little tricks.

Content of multiple-choice tests. Opinions differ with respect to the completeness of the material dealt with in multiple-choice tests. The lower stream of secondary education (LTO) expresses more faith in reading comprehension tests than the higher stream. Still, students clearly miss the opportunity to express their own ideas about the text, particularly for Dutch.

Degree of difficulty of multiple-choice tests. Students do not find multiple-choice tests more difficult than open-question tests. A minority also finds that they receive higher grades on multiple-choice reading comprehension tests than on other tests.

Multiple-choice tests and language skills. The student reactions on the last category of statements show that one aspect students like about multiple-choice tests is the fact that they cannot be judged on language mistakes. They do not show different answer-percentages for the various school subjects. The idea that an exclusive confrontation with multiple-choice questions might influence productive language skills unfavorably meets with much agreement.

9. CONCLUSIONS

On the basis of the research reported here, the overall conclusion may be drawn that in the Netherlands many of the initial fears about multiple-choice testing have turned out not to be based on facts. The Dutch system may have struck the proper balance between multiple-choice and traditional testing, by continuing the use of other testing techniques, side by side with multiple-choice tests. This has minimized the risk of educational 'impoverishment'. It is still necessary, though, to carefully follow the

developments in this area, to prevent an adaptation of the educational system to the testing technique.

FOOTNOTES

1. HAVO: middle stream of secondary education, 5 years.
   LBO : lower stream of secondary education (vocational track), 4 years.

2. In the same survey, students were also asked about their study-habits for multiple-choice and open-question tests for Chemistry and Biology. In contrast with results from earlier research, students did not turn out to pay more attention to details, or learn more by heart when preparing for multiple-choice tests. The answers show more of a tendency towards a more global method for multiple-choice tests: students underscore more, write down less, and learn less by heart.

   For all school subjects most students preferred multiple-choice tests, while open-question tests were said to be more difficult for all school subjects than multiple-choice tests.

REFERENCES

Davis, F.B. (ed.) The literature of research in reading with emphasis on models. New Brunswick (N.J.), Rutgers - The State University, 1971.

Wesdorp, H. (ed.) Studietoetsen en hun effecten op het onderwijs. 's-Gravenhage, Stichting voor Onderzoek van het Onderwijs - Staatsuitgeverij, 1979.

Table 1. Mean test-scores HAVO-English

Socio-economic status	N	Multiple-Choice	Open-Q	Difference-scores
1. unskilled and skilled labor	64	4.74	2.40	2.34
2. lower level employees	108	4.43	2.84	1.59
3. small business-owners	45	4.11	2.28	1.83
4. mid-level employees	44	4.68	3.07	1.61
5. high level employees	69	4.46	2.59	1.87

Table 2. Mean scores and standard-deviations for students of different of different socio-economic status: final examination Dutch (LTO 177)

	N	Multiple-Choice M	Multiple-Choice SD	Open-Questions M	Open-Questions SD
1. Unskilled labor	29	26.69	4.67	36.69	6.70
2. Skilled labor	95	28.74	5.27	36.18	6.62
3. Lower level employees	46	28.48	5.51	37.02	5.69
4. Small business-owners	58	29.52	4.85	36.81	6.68
5. Mid-level and high level empl.	30	29.83	5.09	36.43	7.71

Table 3. Mean scores and standard-deviations for old (1966) and new (1977) essays, by means of global and analytic judgement by jury's of 15 and 16 members, respectively

	GLOBAL		ANALYTIC							
			Total		Content		Style		Conventions	
	Old	New	Old	New	Old	New	Old	New	Old	New
M	6.50	6.60	5.96	6.04	2.64	2.77	2.84	2.79	2.95	2.80
SD	0.68	0.67	0.67	1.02	0.38	0.56	0.44	0.57	0.52	0.61
n	64	64	32	32	32	32	32	32	32	32

Table 4. Time spent on various teaching-methods in 51 'test-schools' and in 51 'non-test-schools' a)

	test-schools		non-test-schools	
type of teaching	n	med.	n	med.
teaching in class	47	3.86	50	3.25
teaching individually	48	2.07	50	2.17
teaching small groups	47	2.00	50	1.62
teacher activities				
teaching	44	2.21	41	0.88
narrating	50	1.63	48	1.29
explaining	51	1.74	48	1.90
demonstrating	49	0.98	45	1.06
individual activity	50	2.17	47	2.00
class discussion	51	1.07	49	1.16
oral quizzing	48	0.95	46	0.67
written quizzing	50	1.07	48	0.90

a) Time spent is expressed on a seven-point scale, ranging from 1 (almost never) to 7 (more than twenty hours per week).

Table 5. Time spent on language curriculum components in 51 test-schools and 51 non-test-schools b)

Curriculum component	Part of test?	Test-schools n	Test-schools med.	Non-test-schools n	Non-test-schools med.
Technical reading	no	51	1.00	50	1.07
Reading comprehension	yes	51	1.36	50	1.23
Essay writing	no	51	0.86	49	0.84
Grammar	optional	48	0.97	48	1.01
Spelling	yes	51	1.25	49	1.13
Style excercises	yes	49	1.14	49	1.16
Listening comprehension	no	49	0.97	48	0.97
Speaking	no	50	0.95	50	1.00

b) Time spent is expressed on a seven-point scale, ranging from 1 (less than one hour per week) to 7 (more than six hours per week).

Table 6. Classification of traditional and objective examinations for Dutch reading comprehension for LTO-schools from 1971 to 1977, by three judges, according to the Davis criteria (in percentages)

Examination	Categories									
Open Question	1	2	3	4	5	6	7	8	9	Not classified
judge I	20	7	20	3	19	17	8	6	0	0
judge II	12	13	16	0	31	10	6	2	0	9
judge III	24	11	19	0	32	6	5	0	0	4
Multiple-choice										
judge I	6	11	6	10	23	26	9	6	3	0
judge II	1	18	7	3	31	16	18	4	3	0
judge III	4	16	10	4	27	20	13	4	2	1

Table 7. Classification of traditional and objective examinations for Dutch reading comprehension for 1976 and 1977 by three judges, according to the Davis criteria (in percentages)

Examination	Categories								
Open Question	1	2	3	4	5	6	7	8	9
judge I	1	3	10	5	41	27	10	2	1
judge II	0	2	13	0	62	15	7	1	1
judge III	0	10	13	0	65	10	1	0	0
Multiple-choice									
judge I	4	9	3	11	25	26	8	9	5
judge II	0	16	5	5	30	21	13	4	5
judge III	4	13	7	6	29	20	11	5	4

Table 8. Use of multiple-choice questions in the classroom and as homework-assignment in the two final years of secondary school, in percentages

	Dutch	French	German	English
Uses multiple-choice questions as teaching aid in teaching subject material	24	20	19	25
Gives students multiple-choice questions in class for practising subject material	52	34	31	38
Gives students multiple-choice questions in class for practising multiple-choice questions	58	40	54	51
Gives students multiple-choice questions as homework for practising subject material	27	24	21	27
Gives students multiple-choice questions as homework for practising multiple-choice questions	31	22	29	29